

**CERTIFICATE OF TRANSMISSION**

I hereby certify that this correspondence (along with any paper referred to as being attached or enclosed) is being submitted *via* the USPTO EFS Filing System on the date shown below to **Mail Stop Appeal Brief - Patents**, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.

Date: November 4, 2008/Jessica Sexton/  
Jessica Sexton**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re patent application of:

Applicant(s): Eric Brill, *et al.*

Examiner: Jude Jean Gilles

Serial No: 10/670,681

Art Unit: 2143

Filing Date: September 25, 2003

Title: SYSTEMS AND METHODS FOR CLIENT-BASED WEB CRAWLING

**Mail Stop Appeal Brief-Patents**  
**Commissioner for Patents**  
**P.O. Box 1450**  
**Alexandria, VA 22313-1450**

---

**REPLY BRIEF**

---

Dear Sir:

Applicants' representative submits this Reply Brief in response to the Examiner's Answer dated September 4, 2008. In the event any fees may be due in connection with this Reply Brief, the Commissioner is authorized to charge such fees to Deposit Account No. 50-1063 [MSFTP475US].

**REMARKS**

Claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112, 114-116 are currently pending and are presently under consideration. Favorable reconsideration of the subject patent application is respectfully requested in view of the comments herein. In particular, the following comments address deficiencies contended in the Examiner's Answer to applicants' Appeal Brief.

**I. Regarding the Rejection of Claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112 and 114-116 Under 35 U.S.C. §103(a)**

The Examiner incorrectly maintains the rejection of claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112 and 114-116 under 35 U.S.C. §103(a) as being unpatentable over Bailey *et al.*, (U.S. 20060167864) in view of Albion *et al.* (U.S. 20040240388). Appellants' representative respectfully requests that this rejection be reversed for at least the following reasons. Bailey *et al.* and Albion *et al.* alone or in combination fail to teach or suggest all features set forth in the subject claims.

Applicants' claimed subject matter relates to data analysis, and systems and methods for obtaining information from a networked system utilizing a distributed web crawler. Information gathered by a server's web crawler is compared to data retrieved by clients of the server to update the crawler's data. In particular, independent claim 1 recites *a data analysis system, comprising: a first component associated with a server of the data analysis system that facilitates generation of a first data set related to web page information obtained via a communication system; and a second component that coordinates a second data set relating to web page information from at least one distributed resource associated with at least a client of the server which interacts with the communication system.* Independent claims 37 further recites *refining the second data set to reflect information obtained from the third data set by adding unknown information to the second data set when new information is received from the distributed source via the third data set.* Independent claims 57, 61, 92 and 113 recite similar features. Bailey *et al.* and Albion *et al.* are silent regarding such novel features.

Bailey discloses a web crawler that locates web pages, the located web pages are scored according to a set of content-based rules. The scores may additionally be based on other criteria, such as content of other web pages of the same web site. On page 18 of the Examiner's Answer,

the Examiner contends that the additional web pages present at a website of a web page located by the crawler program disclosed by Bailey *et al.*, can be submitted by clients or another server. However, Bailey *et al.* discloses that web pages located by a crawler are sent to the product score generator where they are scored, the content of the additional web pages located at the web site hosting the crawler located web page are also evaluated to score the crawler located web page. Thus, the additional web pages are located by the product score generator, from information gathered from the initial web page provided by the web crawler. However, nowhere does Bailey *et al.* disclose that the communication system (internet) from which the web crawler locates web pages, is utilized by the product score generator to locate the additional web pages and update the initially located web page with data from the additional web pages. Rather, Bailey *et al.* discloses that web pages are input to the product score generator from the web crawler, scores are generated for the web page based on evaluation of content (*See* Bailey *et al.* paragraph 0037). Further Bailey *et al.* discloses that the web crawler and the product score generator run on different platforms, however, the cited reference is silent regarding a server client relationship between them. In contrast, the claimed invention provides for a server that hosts a web crawler searching a communication network such as the Internet for other servers hosting web pages, gathers information about these web pages and compiles them for utilizing with a web page search engine (*See* applicants' Fig.1 and Fig.2). The server then sends a representation of this web page information to a client of the server. When the client accesses that particular web page or detects web pages that are unknown to the server, the client compiles changes/status and/or new information about the known and unknown web pages. This information is then transmitted to the server, which utilizes the information to update its original crawler web page data to reflect a new web page or change of contents in a known web page.

Thus, Bailey *et al.* is silent regarding utilizing web page information communicated by a client of the distributed web crawler system to update its original crawler web page data to reflect a new web page or change of contents in a known web page. For example, Bailey *et al.* teaches a conventional web crawler implemented by a server but does not teach or suggest that the web crawler 160 is updated with inputs from the clients 110 (*See* Bailey *et al.* Fig. 1 and paragraph [0037]) Thus Bailey *et al.* does not disclose a distributed web crawler wherein a client updates web pages associated with a server of the distributed system as recited by the subject claims.

Albion *et al.* relates to dynamic assignment of timers in a network transport engine that provides a connection between two applications running on different system interconnected via the network. On page 18 of the Examiner's Answer, the Examiner equates updating of client requests and/or timeout timer lists by the crawler to the feature of *refining a second data set to reflect information obtained from the third data set by adding unknown information to the second data set when new information is received from the distributed source via the third data set*. The second data set disclosed by Albion is the timer list or client request list in the timer logic, the list data is already present in the timer logic, generated by the same crawler that provides the third data set. Nowhere does Albion disclose the second data set is obtained from web pages interactive with a server of a communication system; the third data set is received from at least one distributed resource comprising a client of the server that is interactive with the communication system disclosed by independent claim 37. Rather, the same crawler generates both the data sets. The client request for the timer is also not web page information communicated by a client of the distributed web crawler system as recited by the subject claims. Thus, crawler 204 of Fig. 2 in Albion is a component that manages timers accessed by clients (*See e.g.*, Albion paragraph [0018]), rather than a component that provides unknown information or updates information when changes have occurred in the contents of the web page information as recited in the subject independent claims. Therefore, it is concluded that Albion *et al.* is silent regarding *generating a first data set relating to a second data set obtained from web pages interactive with a server of a communication system; receiving a third data set from at least one distributed resource comprising a client of the server that is interactive with the communication system; the third data set comprising web page related information generated by the distributed resource; refining the second data set to reflect information obtained from the third data set by adding unknown information to the second data set when new information is received from the distributed source via the third data set* as recited by independent claim 37.

In view of at least the foregoing, it is readily apparent that both Bailey *et al.* and Albion *et al.* fail to teach or suggest all limitations of the claimed invention. Accordingly, it is respectfully requested that rejection of independent claims 1, 37, 57, 61, 92 and 113 (and the claims that depend there from) be withdrawn.

**II. Conclusion**

The subject application is believed to be in condition for allowance in view of the above comments. A prompt action to such end is earnestly solicited.

In the event any fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP475US].

Should the Examiner believe a telephone interview would be helpful to expedite favorable prosecution, the Examiner is invited to contact applicants' undersigned representative at the telephone number below.

Respectfully submitted,  
AMIN, TUROCY & CALVIN, LLP

/Himanshu S. Amin/  
Himanshu S. Amin  
Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP  
24<sup>th</sup> Floor, National City Center  
1900 East 9<sup>th</sup> Street  
Cleveland, Ohio 44114  
Telephone: (216) 696-8730  
Facsimile: (216) 696-8731